Commentary

LOWELL A. MARTIN

THE PAPERS in this issue identify the recurring problems of library evaluation: vague objectives, uncertain measures, and half-formulated standards. Before examining these problems, comment must be made on the conception of evaluation—its purpose and nature—that is assumed by the various authors and that limits the total impact of the issue.

EVALUATION AS AN INTEGRAL AND CONTINUING PROCESS

Evaluation is seen in most of the present papers as a fairly distinct and separate activity in library planning and administration, a pause to take stock and assess strengths and weaknesses. At a given point in time a cross-section is extracted and held up to examination. This is like a periodic visit to the doctor, and as such is to be commended. The "doctor" in the case may be an outside surveyor commissioned to give a diagnosis and prescription. In the evaluations properly required for many federally funded projects in recent years, the measurement and assessment responsibility was explicitly given to separate and independent assayers in order to get judgments free of vested interests. A library may embark on a project of self-evaluation on its own, for example examining the collection by means of staff committees or assessing reference service by means of a month-long sample analysis.

Whoever the evaluators may be, the process of evaluation in this more common conception is an activity apart from the normal order of operation and service. When the project is done, and conclusions are in hand, one seeks to apply them, adjusting program or method as the results indicate, and then gets back to the regular business of the day.

One limitation in this approach is the difficulty in jumping from measured conditions to diagnosis of what is wrong and then jumping again to a prescribed course of action. The doctor can make a mistake at either step—and the wise doctor has the patient come back in a day

Lowell A. Martin is Professor, School of Library Service, Columbia University, New York, New York.

or a week or a month, in order to check his analysis and change or adjust his prescription if necessary, thereby setting up a process of feedback and continuous evaluation. Assuming the right and fitting conclusions are drawn from a single, finite evaluation study, they may still not be applied correctly. Their presentations may not have been convincing, or results could be interpreted in different ways, or proposals for action may run into entrenched interests. Many a commissioned study stands unheeded on the shelves because of some combination of these factors. Even if the findings of the evaluation are convincing and are applied, results tend to be singular rather than continuous. The evaluation revealed or documented certain shortcomings, adjustments are made to meet them, and then at some indefinite future time—but perhaps never—another examination will be made to see if the hoped-for improvements actually occurred.

Evaluation conceived as a separate, discrete endeavor, a pause to see where one stands, is likely to lead to a report (after all one has to show that the project is finished), but is less likely to lead to change in the organization. Short of revolution, institutions and bureaucracies do not sharply alter course but at best evolve, give a little here and there, try and try again, in an ongoing process.

Evaluation, to have effect, should be part and parcel of this process. There are occasions when the separate intensive study, either by staff or by outsiders, can lead to action—for example, when a new administrator seeks to set course, or when some problem has become enmeshed in emotion and/or politics and must be freed by concentrated examination. But in most circumstances evaluation and action should be in a constant interchange, if for no other reason than to confirm that previous decisions were wise and effective, more likely to affect recurring adjustment because more often than not human judgment is somewhat short of omniscience.

Back a few decades POSDCORB was the approved formula for administration, the letters of the acronym standing for a neat progression from planning to budgeting. Oddly enough, evaluation, reassessment and appraisal were not included in the formulation. Injection of continuous evaluation and feedback into corporate and institutional decision-making has been one of the notable developments of recent years. Business early established financial controls—after all, its objective is to make money, and this is the measure of its success. Automobile companies, for example, that built in such financial evaluation have flourished while those that failed to do so have not survived, no matter what the quality of their product. Government and institutions have

sought to follow suit—thus PPBS (Planning-Programming-Budgeting System), MBO (Management By Objective) and CIPP (Context, Input, Process, Product). The limited success in this regard in the public sector is due to the recurring problems of evaluation's lack of clear objectives (profits in the case of business) and uncertain measures (sales and dollars). By its standards business makes profits or goes bankrupt; the public agency is more likely to ask for a supplementary appropriation.

Building in evaluation as an integral part of planning and decision-making is not easy. One can mobilize for the one-shot study, or commission it from the outside. Continuing evaluation, on the other hand, calls for (1) a critical and appraising outlook on the part of staff and administrators towards their endeavors, which is not compatible with the faith that sustains many librarians, (2) skill somewhere within the organization in the techniques of measurement, and (3) hard-headed appraisal of results independent of those who have a stake in the success of the enterprise being evaluated.

The closest libraries have come to continuous, integral evaluation is in the systems offices recently set up in a few of the larger university libraries. Even here there are significant limits, for systems analysis is more likely to lead to changes for the sake of efficiency than to changes for effectiveness. We depend now primarily on administrators to set the process of evaluation in motion within the organization, and only a minority have the objectivity and the security to be willing to unlease a force that might well reflect unfavorably on that very administrator's stewardship. Evaluation, carried to its conclusion, can create tension, and only those who know their Mary Follett [1] and the value of conflict in an organization are prepared to make waves.

Examples of continuing and integral evaluation in library projects can be cited. A current Philadelphia inner-city project has an ongoing but independent evaluation staff, reporting by means of rigorous methods, but reporting back monthly and not holding its evidence for some final blast that would pronounce once and for all on the situation from the outside.[2] The Deiches Studies of the Enoch Pratt Free Library [3] have utilized an outside consultant, but in each of the series of studies over ten years—each directed at a problem identified by the Library—the results are incorporated into the organism before the next study is undertaken. By whatever means, it is the immediate feedback, the interchange of action and appraisal, that promises to get systemic results from evaluation.

Beware the separate single-shot evaluation that may be more exor-

cism or expiation than resolve to do better. Espouse built in and continuing evaluation that assesses and redirects daily and monthly and yearly, seeking to adjust the steering wheel frequently rather than taking one turn to get to the top of the mountain.

CLARITY OF OBJECTIVES

The papers in this issue consistently point to definition of the purposes of the agency or project being appraised as the starting point of evaluation. How can one judge the adequacy or effectiveness of any enterprise except against what it is designed to do? We all subscribe to this logic, but there are few fresh statements of purposes and functions of libraries.

Objectives and purposes can be set at several levels from societal and institutional to the library within an organization and to an activity within a library. The less encompassing the level—the more circumscribed the activity being appraised—the easier the definition of objectives and the more conclusive the evaluation. Thus we have the specific examples of relatively satisfying evaluation reported in this volume: interlibrary loan performance as one element in collection appraisal (objective: fill all or most requests received); appraisal of reference service by anonymous shoppers (objective: provide accurate information); the inventory assessment of adult services (objective: readiness to serve). But no one of these indicates whether the loans or information or services contribute to the social ends for which the libraries exist.

Academic, school and special libraries derive their purposes from the agencies they serve. Their task is to define their particular role or roles within the organization, showing how they contribute to institutional goals either directly or in support of other units. This is difficult enough, and more than one company, school, or college library languishes because the task has not been completed or its results not convincingly communicated.

The public library is particularly at sea in this regard, for it seeks to clarify its mission directly at the level of societal goals, not having the guidance of institutional objectives. Lacking institutional charts to steer by, the public library pursues an uncertain course. From its founding it has been pushed by conflicting winds. On the one hand is a professed goal of an educational character, which—if it means anything beyond an approved word—implies selecting among alternative purposes and mounting of collections and programs that contribute to the aims selected. This is not the way the public library builds its program. On

the contrary, it listens to hear what people want and then serves demand: this is called being responsive to the community. This is in the sense that the department store, the liquor store and the candy store are responsive to the community, except that they do not claim to be educational agencies nor do they put in a request for funds from the public purse. It is interesting to note that social scientists, commissioned to examine one or another aspect of the public library, ask first about objectives and when they get vague answers proceed to their own formulations, which often leave librarians uneasy.[4]

Until objectives are clarified and rendered in functional terms, library evaluation will neither characterize the agency for nonlibrarians (governmental officials, educational authorities, the public at large) nor furnish verified judgments on which the library administrator can act. That is the reason that some of the efforts reported in this issue —the user studies mentioned by Monroe, the checklists mentioned by Bond, the reference transaction enumerations mentioned by Weech—are recognized by the authors to be limited and subject to criticism. Whatever degree of validity the evaluation studies of federally financed projects mentioned by Mahar may have derives from assessment within the stated purposes of the grants.

If clarification of objectives is so important for evaluation as well as for other purposes (public relations, for example), why have librarians made so little progress in this regard? It is in part, I am sure, because the exercise is never easy for public agencies; the roles of such agencies are usually a compromise between mission and demand. Also, partly I suspect because librarians recognize that an aura attaches to their agency—each member of the public having his own concept of "library"—and librarians do not want to dispel this aura with hard choices that may please some supporters but would probably alienate others. And finally, I fear, it is because some librarians want to stand on the purity of their motives and the rich humanity of their collections, and do not want to be circumscribed by concrete objectives, as are most other toilers, from automobile salesmen to wearers of the cloth. How else does one account for the fact that librarians as a group, assembled in solemn conference, subscribe to fine objectives which are not carried out back home?

VALID MEASURES

If one knows what an agency seeks to do, the next problem in evaluation is to find measures of what is done. The authors of this volume point to the pitfalls for librarians in taking this step, and search

out the newer efforts to get accurate and relevant statistics about library service.

It is clear from these reports that more attention has been paid to preparation for performance than to performance itself, by practitioners and evaluators alike. We know far more about what goes into a library than what comes out of it. What titles are held? How many librarians are on the staff? What services are provided? This is comparable to an inquiry into schools that examines all aspects of the agency except what children learn. Schools, like libraries, having stressed input rather than output, were ill-prepared for the hardheaded output studies of recent years: the Coleman report,[5] which found only limited correlation between such cherished aims as limited class size, extent of education of teachers and number of volumes in the library, and the performance of students, and the Jencks report,[6] which found limited relation between amount of education and earning capacity in adult life. What if we were to find that a public library which provides multiple copies of best sellers has no more impact in terms of agreed upon objectives than one that expects readers to buy such publications in the marketplace, or if we were to find that a college library that extends its holdings to several hundred thousand volumes influenced the education of students no more than one with a smaller and highly selective collection? Evaluation by input measures serves those who work in an agency, but not those who use the agency, nor those who must pay for it.

Several of the articles in this issue refer to user studies as a growing trend in library evaluation. This is a step in the right direction, for it gets closer to performance and effect; how close it comes, however, is a moot question. People judge a service on the basis of what they expect from it; they have an assumed standard, and it is often modest and tentative. To many—the community resident, the student, the official in business or government—the library is a welcome and somewhat unexpected aid, a largess, and they are grateful for whatever they get. According to Weech's article they are even grateful for information which is inaccurate, for studies which he reports show that librarians believe the information they supply to be accurate, and users accept it as such, when in fact it may be incomplete, out-of-date, or plain wrong. Preconceptions held by users may color their assessments; thus, Mexican-American families have a tradition of not borrowing, particularly from public sources, and scientists, according to some studies, favor small working collections rather than comprehensive libraries. Users are not well qualified to suggest new and additional services from

which they would benefit; one first invents the automobile and television and even Coca-Cola, and makes them available, rather than waiting for people to ask for them. User data are one source of evaluative measures, and should be added to the battery of indicators, but like any measure should be interpreted for what they do not as well as what they do indicate.

Two commonly compiled library measures do deal with output rather than input: circulation statistics and reference counts. It is fashionable to disparage circulation figures. Certainly they have limited validity, for example, in measurement of a special library that provides much of its service by telephone. But in a library that seeks to encourage use of materials, and in which a significant part of the use occurs off the premises, circulation figures are a valid measure of response. Other things being equal, increased circulation figures denote greater impact for whatever the purposes that prompted the acquisition of resources in the first place, and if circulation goes down that impact is diminished no matter what the ingenious justifications that may appear after the fact in annual reports.

Of course circulation is not the sole or the complete measure of a library. In a multipurpose and multiprogram agency there is no one and complete measure of performance, and a search for the magic number only compounds the problem. Surely evaluation of an organization involves more than one computation. Circulation in a public library may account for one-half of what the agency does—and this would seem to be an argument for using and refining it rather than rejecting it. Further, what figures we have show a considerable correlation between circulation and other use figures in the community agencies.

Of course, one does not use circulation statistics from two libraries serving quite different clienteles as a basis for comparison of effectiveness. Every study of adult library use shows a relation between educational level and amount of response. Again, this would seem not to be a reason to reject the measure but rather to refine it so that response with educational level controlled could be determined, a by no means esoteric statistical manipulation. With the relation between clientele characteristics and prospective use known, variable standards can be devised which would show, for example, whether ten items circulated per capita in a community with an educational level close to college graduation represents greater or less response than five books per capita in a community with an average of eight years of education.

Reference and information service, most librarians agree, is of great

importance, and indeed is sometimes held up as the heart of the matter. Despite this I have yet to find a library that has kept meaningful reference statistics over a sustained period that would permit the tracing of trends. From time to time categories for recording are set up and then become blurred; groupings are then simplified and still the record is incomplete. I cannot convince myself that reference service is really so complex and subtle, as compared with what goes on in a school or a hospital, that it cannot be recorded within reasonable limits of accuracy. If administrator and staff were really convinced of the value of such data, and genuinely committed to evaluation of what they do, the brain power and the diligence of librarians is certainly adequate to get the record straight.

The problem is not to find one measure of impact or effect but to identify several that together indicate activity and response in relation to objectives. Indeed it may not be measures as such that we are seeking (the term suggests a yardstick or a thermometer), but rather indicators that objectively reflect reality but that still require judgment for interpretation. The current study at Rutgers University, under the bold title "Measurement of Effectiveness of Public Library Service," should carry us some distance in this direction for one kind of library service. The "document exposure" index proposed in the recent study by Morris Hamburg,[7] while admittedly exploratory, probes toward an indicator that may apply to the various types of libraries.

STANDARDS FOR EVALUATION

Even with objectives clarified and measures in hand, one still needs a criterion, a bench mark, against which to interpret data. A recent issue of *Library Trends* was devoted to library standards, and need not be recapitulated here. But perhaps because of that previous issue, standards are given only limited attention in this volume, and therefore deserve some comment as an element in evaluation.

Actually most standards in the library field have not been designed as aids in evaluation. The more recent statements have been exhortations to newer concepts of library service (library systems in the 1956 statement for public libraries and media centers in the 1960 statement for schools, the former refined in a 1966 document and the latter in the 1969 school media standards).[8] Earlier they were formulations of minimum levels—not "standards" at all if this means measures of quality—and were designed to bring up stragglers among libraries to a kind of tolerable level of mediocrity. In content they have stressed principles—the dicta by which a library should be run in order to

conform to prevailing professional concepts of good service—and have included relatively few measures or indicators. What measures are prescribed in the official standards documents have not been validated: do we know that 1.5 books per capita will meet the needs of a community, or that one librarian for each 250 students in a school can provide full media service? Some of the so-called standards are inadequate on any logical grounds. Compare, for example, the broad and noble objectives set forth at the outset of the public library document, and then consider whether they can be achieved by providing one librarian for every 2,500 people; the crews on the garbage trucks in cities and the number of policemen exceed this ratio many times.

What are the sources of existing standards, the basis on which they have been built? One approach is to determine what exists at a given time and raise the level a little to create a "standard," as though a desirable standard of health were to be not quite as sick as most people. Another approach is to pool the wisdom of the seers, usually meaning the professional seers, and set forth their combined judgment. The library field has used both of these approaches to the present time. Or, as a fresh alternative, one can approach standards as an essential element in evaluation, and go through the laborious but necessary sequence of first making objectives explicit, then establishing measures or indicators that bear on the objectives, and then determining the amount or extent of the indicators necessary to achieve the established objectives. This amount or extent or degree constitutes evaluative standards.

Several library groups are again at work revising existing standards. They can raise the prevailing figures a notch or many notches, depending on their disposition; or they can seek new concepts of service or organization and remake standards around them. If they want to establish measures of library effectiveness, they will have to go through the full evaluative process, all the way to validation of proposed standards in the real world with real people.

Evaluation of an agency occurs either when those responsible for it are enough concerned temporarily to call in evaluators, or enough concerned permanently to institute continuing evaluation as an integral part of planning and decision-making. Without that concern, relatively little evaluation occurs. Libraries face many technical problems of defining measures and recording statistics, but these can be solved—indeed would have been solved by now if the concern were great enough. Now some voices are being raised for more self-

examination, and more funding authorities are asking for evidence of accomplishment, so that some efforts are underway. But among libraries generally little real attention is being paid to the matter. This raises the question as to whether librarians as a group really want an evaluation of their institution.

The response of librarians, assuming it is negative, may not be the decision that prevails. Service agencies, as well as commercial producers, are on trial in a culture that has developed a deep-running scepticism. Schools are being subjected to scrutiny as never before, and universities have gone through trial by fire, not without being burned. Public libraries as yet have not been challenged as much, except by individuals within their own ranks, but the uneasy shifting about of top public library administrators among positions and even their dismissal in some cases may be a symptom. On the public authority side, rather than a challenge being laid down, too often there is dismissal of the request for increased funds. Libraries directly serving the manufacturing and retail and financial complex have had a bit of holiday, while profits have been at peak levels, but the prognosis for company libraries and industrial research libraries may change if the warnings from economists of a downturn in the private sector prove true.

Broad scepticism and questioning come to a focus in the pressures on fiscal authorities, in cities, in schools, in universities. The several types of libraries have felt the financial pinch, and it will get worse before it gets better. They are challenged to prove their worth, and either librarians will come forth with evaluative data to support their case, or officials will assume evaluations that will undermine the case. This accounts for the stance of the present federal administration in relation to libraries: this service is not essential, the administration has said, and the library world has not been able to come up with evidence to the contrary.

Not only whole agencies are being evaluated by the public in general and fiscal authorities in particular, but performance by individual staff members will increasingly be judged. Accountability is a tricky game, but it will be played. One response is to feint and dodge, hoping the questioners will tire and go elsewhere. Or evaluative data on individual performance can be responsibly gathered, and it could even turn out that librarians are relatively productive workers.

For these several reasons—professional, financial, personal—librarians would do well to mount their own evaluative programs. The papers in this volume show that some effort is being made along this line, but they show even more the limited progress that has been

made. Evaluation has been a marginal activity in libraries, engaged in sporadically, often carried out half-heartedly, its results applied reluctantly. The record to date, as reported here, shows libraries to be in a pre-evaluative stage, by turn curious about or perturbed by the next stage, and occasionally reaching forward into it.

References

1. Follett, May P. *Dynamic Administration; The Collected Papers of Mary Parker Follett.* Henry C. Metcalf and L. Urwick, eds. New York, Harper, 1941.

2. Benford, John Q. "Philadelphia Project," *Library Journal*, 96:2041-47, June 15, 1971.

3. The first in the series was *Students and the Pratt Library* (Deiches Fund Studies of Library Service, no. 1). Baltimore, Enoch Pratt Free Library, 1963. The fourth volume in the series, on adult services, is scheduled for publication in early 1974.

4. For example, see: Banfield, Edward C. "Some Alternatives for the Public Library." *In* Ralph W. Conant and Kathleen Molz, eds. *The Metropolitan Library*. Cambridge, Mass., M.I.T. Press, 1972, pp. 89-100; and Hamburg, Morris. *A Systems Analysis of the Library and Information Science Statistical Data System*. Washington, D.C., U.S. Department of Health, Education and Welfare, 1970.

5. Coleman, James S. *Equality of Educational Opportunity*. Washington, D.C., U.S. Department of Health, Education and Welfare, Office of Education, 1966.

6. Jencks, Christopher. *Inequality; A Reassessment of the Effect of Family and Schooling in America*. New York, Basic Books, 1972.

7. Hamburg, *op. cit.*, pp. 251-322.

8. American Library Association. Public Libraries Division. Co-ordinating Committee on Revision of Public Library Standards. *Public Library Service: A Guide to Evaluation, with Minimum Standards*. Chicago, ALA, 1956; American Association of School Librarians. School Library Standards Committee. *Standards for School Library Programs*. Chicago, ALA, 1960; American Library Association. Public Library Association. Standards Committee. *Minimum Standards for Public Library Systems, 1966*. Chicago, ALA, 1967; and American Association of School Librarians and the Department of Audiovisual Instruction of the National Education Association. *Standards for School Media Programs*. Chicago, ALA, 1969.

This Page Intentionally Left Blank